MULTIMEDIA UNIVERSITY®

STUDENT ID NO

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2022/2023 (TERM ID 2210)

HPB3011 –BIOINFORMATICS ALGORITHMS II

(All sections / Groups)

13 FEBRUARY 2023
9.00 am – 11.00 am
(2 hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of 4 pages, including this cover page.
2. You are required to attempt all questions. All questions carry equal marks (10).
3. Write all your answers in the Answer Booklet provided.
4. You may use a calculator.

Question 1

a) What are the TWO disadvantages and advantages of non-hierachichal clustering? [2 marks]

b) You are supposed to perform transcriptomics analysis of cancer sample before and after immunotherapy. One of your genes of interest contains six isoforms. In addition, it has three paralogous genes. Should you use single-end or paired-end read for RNA-seq experiment? Explain your answer briefly. [2 marks]

c) "*A student designs an experiment to compare gene expression between two samples of yeast grown under hot and cold conditions. The student spotted DNA from hot and cold yeast, total yeast DNA is labelled and hybridized to the slide. However, the student observed that the data does not make sense and difficult to analyse.*"

The student made mistakes in designing the experiment. Based on the scenario above, propose a proper experiment design. [2 marks]

d) Why gene and transcript expressions could not be conclusively inferred as the same data? [1 mark]

e) Single dataset was used in two assemblies with different parameters. The assemblies resulted in **transcriptome** with the N50 of 10000 and 35000 respectively. Which of the assemblies is considered as good quality? Justify your answer. [2 marks]

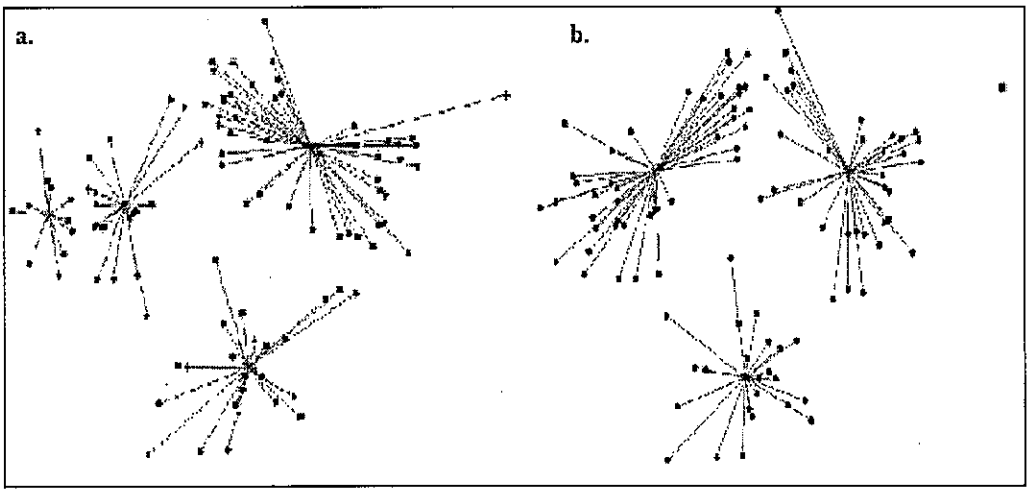
f) Give TWO advantages of RNA-seq compared to DNA microarray for the study of transcriptomics. [1 mark]

Continued .....

Question 2

a) List any FOUR factors for suitability of machine learning in bioinformatics analysis. [2 marks]

b)



The figure above shows two k-means clustering results from single dataset. What is the possible factor contributing to the observation? Suggest two ways to terminate the iteration in k-means clustering? [2 marks]

c) Using support vector machine, describe the outcome if the feature space is too huge. [1 mark]

d) What is the difference between recurrent and feed forward artificial neural network? [1 mark]

e) Given the following probabilities, use the Bayes approach to find the probability of having viral infection if the test is positive – P(Infection |Test+). [2 marks]

	Infection	NO Infection
Test +	0.85	0.03
Test -	0.15	0.97

overall probability of diabetes = 0.01

f) Explain the differences:
   i. supervised learning and unsupervised learning
   ii. training data and testing data. [2 marks]

Continued .....

Question 3

a) In secondary protein structure prediction, CChou-Fasman algorithm is known as a knowledge based algorithm. Explain why. [1 mark]

b) In PHD secondary structure prediction server, what are the TWO artificial neural levels? Explain the differences. [2 marks]

c) What is Levinthal paradox? [2 marks]

d) A graduate student intends to conduct analysis on protein sequence isolated from MERS-CoV. The student wishes to analyze the functional and evolutionary aspects of the protein. Suggest two databases for the purpose. Explain the suitability of the databases. [2 marks]

e) What are the THREE categories of protein secondary/tertiary structure prediction methods? Briefly explain the principles for each method. [3 marks]

Question 4

a) Given an RNA sequence "CCCUUUUAGG". Use Nussinov algorithm to predict the secondary structure. [3 marks]

b) What is the disadvantage of Nussinov over mutual information and Co-variance algorithms in RNA secondary structure prediction? [1 mark]

c) One of the functions of pseudoknot is to serve as chaperone. How is the chaperone function different than ubiquitin? [1 mark]

d) What are the basic principles for RNA secondary structure prediction based on free energy minimization? [2 marks]

e) What are the TWO features of RNA that complicates the structural prediction? Explain. [2 marks]

f) Name ONE energy minimization algorithm in RNA structure prediction? [1 mark]

END OF PAPER